Review of Marketing Science Working Papers

Volume 1, *Issue* 2 2002 *Working Paper* 1

The Impact of the Internet on Information Search for Automobiles

Myung-Soo Lee* Brian T. Ratchford†

Debrabrata Talukdar‡

*City University of New York

†University of Maryland

‡State University of New York at Buffalo

Review of Marketing Science Working Papers is produced by The Berkeley Electronic Press (bepress). http://www.bepress.com/roms

The Impact of the Internet on Information Search for Automobiles

Brian T. Ratchford
R.H. Smith School of Business
University of Maryland
bratchfo@rhsmith.umd.edu

Myung-Soo Lee
Zicklin School of Business
Baruch College
City University of New York
mslbb@cunyvm.cuny.edu

Debrabrata Talukdar
School of Management
State University of New York at Buffalo
dtalukda@ascu.buffalo.edu

Brian T. Ratchford is the Pepsico Foundation Professor of Consumer Research at the R.H. Smith School of Business, University of Maryland. His research centers on the role of information and retailers on consumer choice processes, and on modeling retail markets. He is Editor of *Marketing Science*, and a member of the editorial boards of *Journal of Marketing Research*, *Journal of Consumer Research*, and *Journal of Retailing*.

Myung-Soo Lee is associate professor of marketing at the Zicklin School of Business, Baruch College, City University of New York, and a mentor in the Lawrence N. Field Center for Entrepreneurship and Small Business at the Zicklin School. He received a Ph.D. in marketing from the State University of New York at Buffalo. His articles have appeared in *International Journal of Forecasting*, *Journal of Business and Entrepreneurship*, *Journal of Advertising*, *Journal of the Academy of Marketing Science*, and several other journals.

Debrabrata Talukdar is assistant professor of marketing at the School of Management, State University of New York at Buffalo. He received his Ph.D. in marketing from the University of Rochester. Before returning to academia, he worked on international environmental policy issues at the World Bank. His research interests include consumers' information search behavior, new product diffusion, and the effect of marketing on the welfare of society.

Corresponding author: Brian T. Ratchford, phone 301-405-2202, fax 301-405-0146.

Abstract

While the Internet has apparently become a major source of information about many items, including real estate, travel, computer equipment, books, records, and automobiles, little is known about how it is used, and about its impact on search behavior. The objective of this study is to shed light on this question by studying the use of the Internet as an information source for one good, automobiles.

In order to accomplish our objective, we need to construct a model of the choice of information sources. We liken the time spent with each alternative source to the choice of factors of production. In our case, different sources are combined to produce current information, which, when combined with prior information, produces gains to information, defined as the difference in monetary value between the best choice given complete information and a random choice. Assuming that consumers choose the optimal amount of time to allocate to each source, given their time costs, we derive first-order conditions that lead to expressions for the share of time spent with each source and total search time. These expressions become the basis for our empirical analysis.

Consistent with past literature and our data, we define four major categories of sources of information on automobiles. These are interpersonal sources such as friends and relatives, non-advocate or neutral sources of print information such as *Consumer Reports*, dealer/manufacturer sources, and the Internet.

We use data on information source use obtained from recent new car buyers in one market in 1999, after the Internet became established, and compare this with data obtained from a survey of recent car buyers obtained in 1989 in the same market, before the Internet was introduced. The two data sets allow us to create baseline estimates of what buyers with a

particular set of characteristics would have done in the absence of the Internet, and then to determine how the presence of the Internet is associated with deviations from this baseline.

Our descriptive results indicate that about 40 percent of our sample used the Internet as a source in their purchase decision, which is very consistent with other evidence on Internet use in 1999, the time of our survey. The share on Internet use in total search is significantly related to younger age, higher education, higher wage, dissatisfaction with the dealer of the previous car, and no decision on which manufacturer to buy from prior to the search.

A testable prediction of our model is that the Internet should draw proportionally from each other source. Since dealer/manufacturer sources get the biggest share of time in the absence of the Internet, this implies that the Internet has the greatest absolute impact on time with these sources. Tests of the proportional draw assumption indicate that this hypothesis cannot be rejected, and that the assumption provides a good approximation for our data.

Our analysis of total search indicates that total search increases with factors that determine gains to search, such as price level of the chosen model, and decreases with prior information and efficiency at search. Specifically, total search decreases with a prior decision on a manufacturer or dealer to buy from, satisfaction with the previous dealer, and age. Search also decreases with wages, a measure of time costs.

Putting all of the pieces of our model together allows us to estimate the effects of the Internet and other consumer characteristics on search, and on gains to search. We simulate overall behavior, and that of Internet users and certain other groups, both with and without the Internet. Our simulations indicate that the Internet both reduces total search time, and leads to improved purchases. These effects are largest for younger, more educated consumers,

particularly those that were dissatisfied with their previous dealer. Conversely, the Internet has very little value for older consumers, and younger consumers with high school or less education.

Our simulations also indicate that the Internet leads to a considerable reduction in time spent with dealer/manufacturer sources, most of which is time spent at the dealership requiring the presence of a sales person. This suggests that, by making it easier for consumers to be better informed when they arrive at the dealer and hence cutting down on the demand for sales person time, the Internet can lead to a substantial improvement in the efficiency of dealer operations. Thus the Internet is likely to benefit dealers as well as consumers.

Keywords: Internet, search, information, automobiles

The Impact of the Internet on Information Search for Automobiles

Introduction

There are currently very few countries and territories left in the entire world that do not have at least one host computer connected to the Internet – the massive network of interconnected packet-switched computer networks (Rutkowski 1999). The number of Internet users worldwide has increased exponentially from about 40 million in 1996 to nearly 300 million in 2000, with 108 million of them in the US alone (Newsbytes News Network 2000). The future growth trend looks equally impressive with a projected 1 billion Internet users by 2005. Its phenomenal growth and worldwide scope as a communication medium have even prompted some to compare the Internet as "the most important innovation since the development of the printing press" (Hoffman 2000).

People are increasingly using the Internet not only as a major source of general news information but also to search for specific product information on a variety of goods and services like real estate, automobiles, travel, education, computer equipment, books and records. Because of its obvious relationship to brand and store choice, understanding consumers' product information search behavior has always been critical to firms' strategic decision making in terms of communication media choice. So, with the Internet emerging as an increasingly popular information source for consumers, it has become important to understand who uses the Internet as a source, how it is used, and its impact on consumers' use of conventional sources of product related information. The answer to these questions is particularly relevant as businesses debate how best to allocate billions of advertising dollars between the Internet and other conventional communication media (Dreazen 1999).

Unfortunately, as Parker and Plank (2000) point out, there has been "little information published about the Internet as an information source and whether or how media habits are undergoing changes due to its rapid growth." With some exceptions to be discussed later, the information mostly consists of a few survey studies in the business press. It is clear from these studies that the use of the Internet has grown rapidly, and has become widespread. For example, J.D. Power and Associates (2000) found that 54 percent of new car buyers used the Internet in 2000, up from 40 percent in 1999. However, studies often reach contradictory conclusions about the impact of the Internet on the use of traditional media like the print and TV (Consoli 2000; Gablehouse 2000). There is a need for study of the use of the Internet as an information source, of impact of Internet use on the employment of other sources, and of the effect of the Internet on total search effort and consumer welfare.

The objective of this paper is to use a natural field experiment for a systematic empirical investigation into the determinants of the use of the Internet as an information source for automobiles, and into the impact of the Internet on consumers' use of conventional product information sources, and on information search in general. Specifically, we use product information search data collected from a sample of new car buyers in an identical field setting before and after the introduction of the Internet. We use these data to determine who uses the Internet as a source, describe their purpose of this use, and estimate the impact of the Internet on the use of other sources, and on total search. We also derive estimates of the impact of the Internet on total search, and on consumer welfare.

While there are numerous theories and models of total search effort (e.g., Moorthy, Ratchford and Talukdar 1997), theoretical frameworks for studying the choice of information sources are not well developed (Hauser, Urban and Weinberg (1993) is an exception). Thus we

had to develop a theoretical framework for studying the choice of the Internet as an information source. In this paper, we build a formal model of the choice of information sources and of total search for information in a cost-benefit framework. This model extends and makes operational the general framework presented in Ratchford, Talukdar and Lee (2001). A testable assumption of the model developed in the present paper is that the share of time spent with the Internet is drawn proportionally from the other sources that we consider. Our empirical evidence, which compares information source use before and after the introduction of the Internet, is consistent with this assumption. In our empirical analysis we find that the share of search time spent with the Internet is inversely related to age, and positively related to education and dissatisfaction with the dealer of the previous car. We also model the determinants of total search effort, and empirically find that factors related to prior information, such as a decision about what manufacturer or dealer to buy from prior to the search, are inversely related to time spent searching. Based on our model, we estimate the net monetary benefit of the Internet to different buyer groups. The Internet provides a considerable benefit to younger, more educated, consumers.

We present a brief review of the academic literature on Internet use and search for automobiles in the next section. After this we present our theoretical framework. We go on to describe our data, and present descriptive information on Internet use that is derived from our data. We then study the shares of time spent with the Internet and other sources, and test whether the Internet had a disproportional effect on the use of other sources. We go on to analyze total search behavior across all sources, and the impact of the Internet on this behavior. Finally we present our estimates of gains due to the Internet.

Literature Review

While it is rapidly expanding, the literature related to the use of the Internet as an information source is still limited. The pioneering article of Alba, et al. (1997) discusses the potential impact of electronic commerce on consumers, retailers and manufacturers. However, this article does not contain a formal model of the choice of the Internet as an information source. Bakos (1997) does present a formal search model that considers the impact of the Internet on equilibrium outcomes. However, the basic focus of the Bakos paper is on market outcomes, not individual differences in search or choice of different sources. The present study focuses on how the latter varies across individuals.

While the above articles imply that the Internet will enhance competition, Lal and Sarvary (1999) and Lynch and Ariely (2000) show that this is not necessarily the case. Lal and Sarvary (1999) present a theoretical model in which Internet can decrease search by lowering the cost of buying a preferred item relative to searching in the store. Lynch and Ariely (2000) present a simulated web search environment for wine. They show that designing the web environment to make quality comparison easy decreases price sensitivity for unique items, giving them more monopoly power. Conversely designing the web environment to facilitate comparison makes common items more price sensitive, but lowers demand for the unique items.

Academic studies of the impact of the Internet on the use of other sources are few. Parker and Plank (2000) recently conducted a survey study of the media habits of students. This study suffers from several limitations. First, it used a sample that was relatively small (204) and was restricted to only undergraduate business students. Also, the design of the study provides only a “snap-shot” of current media habits, rather than any insight into the impact of the Internet in terms of *change* in media habits. Finally, this study does not fully control for demographic variables to explain differences in media habits across individuals.

Ratchford, Talukdar, and Lee (2001) present a general model of the choice of the Internet and other information sources. In this model the use of specific types of sources depends on types of attributes that are salient, prior information, skill at using each source, ease of accessing a source, and income. Using this model, a number of general propositions about the use of the Internet are developed. The model in Ratchford, Talukdar, and Lee (2001) is a general theoretical model that cannot be readily employed empirically. The model in the present paper is an extension of their framework to the specific case of search for automobiles, and contains specific functional forms for prior information and the productivity of sources that can be parameterized. The resulting model of the choice of sources is estimated in this paper. While Ratchford, Talukdar, and Lee (2001) present some univariate contrasts between users and non-users of the Internet in searching for automobiles, they do not present a complete multivariate analysis of the type presented in the present study.

Two recent working papers consider the use of the Internet in search for automobiles, and present results that are complementary to those presented in this study. Klein and Ford (2001) collected survey data using the Internet on search for automobiles from 369 respondents who were either recent buyers or current shoppers. A key finding of their study is that on-line use appears to have three dimensions: (1.) manufacturer/dealer sources, (2.) buying services and other third parties, and (3.) bulletin board/chat sources. These dimensions each have off-line counterparts: (1.) dealer visits/advertisements; (2.) *Consumer Reports* and other third-party sources of print information; (3.) interpersonal contact with friends/relatives. Since the data for the Klein and Ford (2001) study were gathered on the Internet, a drawback of this study is that it does not contain information about the information source use of car buyers who do not use the Internet.

Morton, Zettelmeyer and Risso (2000) study the effect of Internet referral services on automobile dealer pricing in California. They find that an average customer of Autobytel.com does, in fact, pay less for a given car than a customer who does not use this referral service. This is evidence that Internet use, or at least the use of Internet referral services, leads to better buys for automobiles.

Though consumer choices of information sources have important implications for manufacturer and dealer advertising decisions and marketing efforts, there is surprisingly little published evidence about consumer use of information sources in search for automobiles. Nor is there much evidence about the use of information sources in general. Furse, Punj and Stewart (1984) present a typology of source use for automobiles that finds clusters of consumers with different patterns; Klein and Ford (2001) replicate this approach for their sample of Internet users. One of the interesting findings of both studies is that there is a large segment of consumers that does not search much at all. Both of these studies are primarily empirical and do not offer a theory of the choice of sources. In contrast, Hauser, Urban and Weinberg (1993) develop and estimate a model of the sequence of choice of sources. Their model was applied in the context of a laboratory simulation of the information search process, and cannot be directly applied in this study since we only have survey data on total time spent with each source. We construct a model that is geared to this type of data.

Theoretical Framework

A consumer's process of information search and acquisition can be conceptualized as a production process in which the consumer seeks to maximize the difference between the utility gain and cost of search. Information sources can usefully be identified as inputs to this production process in which time spent with each source leads to increased information, and

ultimately a better decision. One source of information is memory or prior knowledge, which is a control in our study, which focuses on the choice of external sources of information.

Based on previous literature and the data available for empirical work, we define the following categories of external of sources of information for our study:[1]

- Dealer/manufacturer-related, defined as retailer/salesman/brochures/ads/test drives/time to get to dealer. These are the primary sources of information on attributes that can only be ascertained by direct inspection (how the car feels in comfort and handling), as well as the place where price negotiation takes place. Because of the nature of our data we cannot easily distinguish activities aimed at information gathering from the dealer from those related to completing the sale. Consequently we count both sets of activities as search.
- Friends/relatives, which are distinguished from other non-dealer sources because information is conveyed by word-of-mouth.
- Non-advocate (neutral) impersonal sources in print media, defined as *Consumer Reports*, and other ratings or price information not sponsored by the dealer or manufacturer.
- Internet, defined as any information conveyed on the Internet, including ratings/price data/ads/manufacturer data/dealer data/on-line chat/the Internet version of *Consumer Reports*.

We construct a model of the consumer's choice of these sources when time is the costly resource used in accessing them. We model the consumer's decision to allocate a total amount of time across sources as follows. Given her opportunity cost of time, the consumer decides how much time to spend on searching each of the sources listed above to maximize the net benefit of

[1] Beatty and Smith (1987) classify sources as media, retail, interpersonal, neutral. Our classification is consistent with this except that media and retailer are combined. This is because we could not separate retailer and manufacturer advertising in our data, nor could we readily differentiate between media advertising and advertising at the point-of-sale, such as brochures.

search. Specifically, let *g* be the difference between the consumer's expected utility of a choice in the focal product class under complete information, and the consumer's expected utility under no information (random choice). Let *I* be a measure of the worth of the consumer's stock of information, which varies between $I = 0$ (if the consumer has no information) and $I = 1$ (if the consumer has complete information). The consumer's gain from information is then gI, which = 0 if $I = 0$ and the consumer makes a random choice, and equals *g* if $I = 1$ and the consumer makes a fully informed choice.

The consumer's stock of information at any time *t* can be expressed as $S + F$, where *S* is prior information and where *F* is the amount of information added to the stock between the time the search commences and the purchase is made. There must be diminishing returns to information, e.g., the function relating *I* to $S+F$ must exhibit diminishing returns. An appropriate function that has *I* bounded between 0 and 1 is $I = 1 - e^{-(S+F)}$. The information produced in the search, *F*, is produced by allocating $t_1,\ldots,t_n$ units of time to *n* different sources: $F = F(t_1,\ldots,t_n)$. A function that reflects diminishing returns to each source is: $F = \sum_j a_j \ln(t_j)$, where the subscript *j* refers to source, we assume $a_j > 0$ if a source is used, 0 if it is not, and that the optimal *F* is positive. For simplicity we assume that the costs of search are exclusively time costs, which would seem to be a reasonable approximation for automobiles, and that each unit of time devoted to a given information source costs *w*. Consequently the consumer's cost of search is $C = w(\sum_j t_j)$.

This leads to a definition of the consumer's maximization problem:

$$\begin{aligned} MaxB &= gI - w\left(\sum_j t_j\right) = g\left(1 - e^{-(S+F)}\right) - w\left(\sum_j t_j\right) \\ &= g\left(1 - e^{-\left(S + \sum_j a_j \ln(t_j)\right)}\right) - w\left(\sum_j t_j\right). \end{aligned} \tag{1.}$$

Taking first-order conditions, we get relationships of the form:

(2.) $$\frac{\partial B}{\partial t_i} = g\left(e^{-(S+F)}\right)\frac{a_i}{t_i} = w \text{ for all } i.$$

The right side of Equation 1 is seen to be constant for all *i*, so that the marginal effects of all of the sources must be equal at the optimum. For the model in Equations 1 and 2 this implies that the relative shares of each source are equal to the relative values of the coefficients a_i and a_j:

(3.) $$\frac{t_i}{t_j} = \frac{a_i}{a_j} \quad \text{or} \quad \frac{t_i}{\sum_j t_j} = \frac{a_i}{\sum_j a_j}.$$

This implies that the share of use of each source identifies the relative values of the coefficients a_i and a_j in our model. Since shares of use of each source identify their relative productivity holding constant gains to information, costs of information and prior information, studying the determinants of cross sectional variation in these shares will be a key to determining differences across consumers in their ability to use different sources, including the Internet.

Determinants of Share of Each Source in Search

In general $a_i/\sum_j a_j$ will vary cross-sectionally with the ability of that source to convey information on the attributes most critical to increasing *I*, and with the consumer's ability to use a source effectively. Thus if the cross sectional variation in the ability of a source to convey information is given by a vector of factors *SI*, and if the cross sectional variation in the ability to use a source is given by a vector of factors, *SU*, then the share of use of source *i* by consumer k varies cross-sectionally as:

(4.) $$s_{ik} = \frac{t_{ik}}{\sum_j t_{jk}} = \frac{a_{ik}}{\sum_j a_{jk}} = \beta_{1i} SI_k + \beta_{2i} SU_k + \varepsilon_{ik},$$

where ε_{ik} is a random error term measuring the impact of unobserved consumer-specific factors on the use of source *i*.

Since there is not much evidence about what determines *SI* and *SU* for specific sources, our expectations about cross-sectional variation in source use must be somewhat speculative. As a first cut, we might expect the relative efficiency of the different sources to vary cross-sectionally as follows:

- Internet (a_{nk}) – Since younger, more educated persons are to be computer literate, and therefore use the Internet more effectively, ability to use the Internet as a source should vary positively with education, inversely with age. To the extent that the Internet is used as a source of information about manufacturers, its share of use should be inversely related to a decision to buy from a particular manufacturer prior to the search. To the extent that the Internet is used for information about dealers and price, its share should be negatively related to satisfaction with the previous dealer. The share of the Internet might be related to the respondent's wage for several reasons: the Internet may be a time saver and the wage a measure of the cost of time; higher wages might be associated with higher income and cars that require more information; higher wage might reflect skills related to Internet use.
- Dealer-related (a_{dk}) – Since all consumers have to use the dealer for price negotiation and actually purchasing the car, the share of time spent with the dealer in the purchase process should vary inversely with the value of information from other sources (dealer share will be high when a_{jk} for other sources is low). This would tend to take place when satisfaction with previous car and dealer are high, a choice of manufacturer has been made prior to the search, and when consumers are older since older buyers tend to be brand loyal and search less. Share of the dealer might also be higher when price

negotiation is a more important part of the search process, which would tend to be true for higher priced cars.

- Friends/relatives (a_{fk}) – Evidence from past research indicates that inexperienced, single, less educated, lower income, female buyers tend to rely heavily on the advice of others in their car purchase (Furse, Punj and Stewart 1984). Consequently we expect the share of friends/relatives in the search process to be inversely related to number of cars bought in the past 10 years, age, education and income, positively related to being female, single and buying one's first car. These inexperienced buyers should also be more likely to buy less expensive cars.
- Non advocate sources in print media (a_{pk}) - These sources of information provide test results and other factual information of the type readily conveyed over the Internet, and major sources of this type of information such as *Consumer Reports* and *Edmonds* are also available on the Internet. One might expect these sources to appeal to educated buyers, but not to be so skewed toward younger buyers, who are more likely to be computer literate and use the Internet instead. Since these sources appear to focus relatively more on model characteristics than price in comparison to the Internet, one might expect them to have relatively more appeal to those who are interested in comparing models. These might be men, who tend to be more interested in cars, and those with an incentive to search for model information, such as those who are dissatisfied with their previous car.

Determinants of Total Search

After taking logs and simplifying, Equation 2 can be written as:

(5.) $$\ln t_i = \ln(a_i) + \ln g - S - \ln w - \sum_j a_j \ln t_j$$

We can eliminate the t_j terms by substituting $t_j = (t_i / a_i) a_j$, which follows from Equation 3, into Equation 5. Performing this substitution and simplifying gives:

(6.) $$\ln t_i = \ln(a_i) + \frac{1}{A+1}\left(\ln g - S - \ln w - \sum_j a_j \ln a_j\right),$$

where $A = \sum_j a_j$, and $a_i > 0$.[2] Thus the time spent with a given source depends on its own productivity, a_i, and with overall incentives to search as given by *g*, *S* and *w*. As one would expect, time with any source increases with possible gains to search (*g*), decreases with prior information, *S*, and with time costs, *w*. The effect of the efficiency parameters *a* on search is complex, and the sign of $\partial \ln t_i / \partial a_j$ is ambiguous. The unobservable term $\sum_j a_j \ln a_j$ can be eliminated from Equation 5 by using $a_j = A s_j$ and $a_j = a_i (s_j / s_i)$. This leads to:

(7.) $$\ln t_i = \frac{1}{A+1}\left(\ln a_i + \ln g - S - \ln w - A \sum_j s_j \ln(s_j / s_i)\right),$$

Noting that Equation 3 also implies $t_i = (\sum_j t_j / \sum_j a_j) a_i$ and $t_j = (\sum_j t_j / \sum_j a_j) a_j$, we can also derive an expression for total search across sources by substituting these terms into Equation 4, and using $a_j = A s_j$:

(8.) $$\ln T = \frac{1}{A+1}\left(\ln(A) + \ln g - S - \ln w - A \sum_j s_j \ln s_j\right),$$

where $T = \sum_j t_j$. Thus total search is related to gains (*g*), prior information (*S*), time costs (*w*), efficiency (*A*), and the distribution of s_j, in the same way as the time spent with a particular source. The shares of use of each source are exogenous to Equation 8 since they are completely determined by the *a* parameters.

[2] If $a_i = 0$ the source is not used.

Cross-sectionally *g* varies with factors that affect the dispersion in the value of cars across consumers. Some of these might be demand shifters (*DS*) such as price paid, stage in life cycle as revealed by age. Thus for consumer *k* we have: $g_k = \alpha_1 + \gamma_1 DS_k + v_{1k}$.

Cross-sectionally *S* varies with factors related to prior information. One would be a pre-decision to buy from a particular manufacturer or dealer (*PD*); another would be a vector of factors related to experience that would provide information on the best buy on the current search (*EX*). These factors would be satisfaction with previous car or dealer, age, since older consumers may have a better idea of what they like, experience at buying cars. Thus for consumer *k* we have: $S_k = \alpha_2 + \gamma_2 PD_k + \gamma_3 EX_k + \nu_{2k}$.

A crude measure of time costs (w_k) is the respondent's wage or income. If *A* were likely to be constant across consumers, Equation 8 could be estimated by linear regression. However this may not be the case since *A* should vary cross-sectionally with factors related to efficiency at search (*EF*). Thus we have: $A_k = \alpha_3 + \gamma_4 EF_k + \nu_{3k}$. Accordingly cross sectional variation in search can be described by substituting terms for g_k, S_k, w_k, A_k into Equation 8. If Equation 8 can be identified, all of the parameters relating a consumer's search effort to gains to search are identified

Effect of Internet on Search

From Equations 3 and 5, the share of the Internet in total time identifies its efficiency parameter relative to other sources. Thus:

(9.) $$s_{nk} = \frac{t_{nk}}{\sum_{j \neq n} t_{jk} + t_{nk}} = \frac{a_{nk}}{\sum_{j \neq n} a_{jk} + a_{nk}}, \quad \frac{s_{nk}}{1 - s_{nk}} = \frac{a_{nk}}{A_k}.$$

Moreover, because our model assumes IIA, the share of any non-Internet source among sources other than the Internet should be the same before and after the introduction of the Internet:

(10.) $$s_{jk1} = \frac{s_{jk}}{1 - s_{nk}} = \frac{a_{jk}}{A_k} = s_{jk0}.$$

This is a testable implication of our model, which will be violated if the Internet takes a disproportionate share away from any other source. We could test this directly if we had pre- and post-Internet observations. While we did not have the panel data needed for this purpose, we did have a pre-Internet survey from which we could estimate a regression of the form of Equation 5. Using this we can get an estimate of pre-Internet behavior $\hat{s}_{jk0}$ for respondents to a post-Internet survey. Under the null hypothesis of IIA,

(11.) $$s_{jk1} - \hat{s}_{jk0} = e_{jk0},$$

where e_{jk0} is a random error term. If the null is violated there should be a systematic relationship between $s_{jk1} - \hat{s}_{jk0}$ and the share of Internet use.

As can be seen from Equation 8, after the introduction of the Internet time spent searching for any consumer *k* can be expressed as:

(12.) $$\ln T_{1k} = \frac{\left(\ln\left(A_{0k} + a_{nk}\right) + \ln g_k - S_k - \ln w_k - \left(A_{0k} + a_{nk}\right)\left(\sum_j s_{1j} \ln s_{1j} + s_{nk} \ln s_{nk}\right)\right)}{A_{0k} + 1 + a_{nk}},$$

where 0 refers to pre-Internet, and 1 refers to post-Internet. The *nk* terms in Equation 12 capture efficiency gains for the consumer due to the introduction of the Internet. From Equation 9, $a_{nk} = A_{0k}\left(s_{nk} / \left(1 - s_{nk}\right)\right)$, which can be substituted into Equation 12 to express it in terms of observed shares. From Equation 8, time spent searching by the same consumer *k* before the introduction of the Internet would be:

(13.) $$\ln T_{0k} = \frac{1}{A_{0k} + 1}\left(\ln A_{0k} + \ln g_k - S_k - \ln w_k - A_{0k} \sum_j s_{0j} \ln s_{0j}\right).$$

If we subtract Equation 12 from Equation 11 we get a complex nonlinear expression in Internet shares. However, if the expression for the Internet share from Equation 9 is substituted into Equation 12 we can estimate the parameters in the system from a pooled version of Equations 12 and 13 which is estimated by nonlinear least squares. Given this equation, and the share equations, all of the parameters in Equation 1 will be identified, and we can make statements about the impact of the Internet on specific individuals.

The Data

Data used in this study came from two independent mail surveys conducted in ten years apart, one in 1990 and another in 2000. The first surveys were sent to a list of 3000 buyers of new cars in the Buffalo, N.Y. area (specifically ZIP codes with first digits 140, 141, 142) in July and August 1989. The list, which was supplied by R.L. Polk, was a systematic random sample of these buyers. A total of 901 respondents were obtained from three mailings conducted in February 1990; due to missing data, the number of cases available for the various analyses varies slightly with less than 901 usable sample sizes.

The second data collection followed exactly the same procedures to make sure that we have two sets of comparable data. The second sample was drawn from the same Buffalo, N.Y. ZIP code area with the same number of 3000 new car buyers during the months of July and August 1999.[3] The list of new car buyers was also purchased from R.L. Polk. A total of 886 usable surveys were obtained from the three mailing made in February 2000. Again the actual cases for each analysis would vary slightly due to missing variables. The timings of the two survey administrations ensure that consumers' behavioral responses are contingent on two distinct situations – with or without the presence of the Internet as a new information source. The

[3] Respondents in both samples are actual purchasers, and those who lease cars are excluded. Strictly speaking, therefore, our results apply only to the population of new car buyers.

two surveys together thus essentially constitute a case of natural experimental field study with the impact of Internet being an experimental variable.

A problem with the survey, which is shared with virtually all studies of search behavior (Moorthy, Ratchford and Talukdar 1997 is an exception), is that several months elapsed between the purchase and data collection, taxing respondents' ability to recall their search. However, automobiles are a major purchase, making it likely that respondents will be able to recall their behavior. Our results are broadly consistent with previous studies of search for automobiles, indicating that our study does not pose any special problems. Moreover the impact of forgetting should be the same for both of our surveys, and therefore cancel out in our analysis. Because of the difficulty of identifying buyers or potential buyers ahead of time, the post-purchase procedure that we employed is really the only feasible method for conducting large-scale surveys of search behavior.

Variables

Consistent with our theory, our dependent measures are self-reports of time spent searching in hours for the sources of information noted above. Total time spent is the sum over all sources of these measures. To check responses, this total was compared with another survey question that sought an estimate of total time. Consistent with past studies, responses showing a discrepancy of more than 50 hours were discarded; there were only a few of these. The correlation between the alternative measures was .84. While we also had data on the respondent's self report of the spouse's search, it was evident that much of this amounted to their going together to the dealer. We decided to limit this study to the respondent's search. Experiments with analyses that included the spouse's search indicated that excluding the spouse did not have any appreciable effect on the results.

Our independent variables consist of measures of amount and content of experience, demographics, prior information, time costs, and price paid. Amount of experience is measured by number of new cars bought in past 10 years; age; years lived in area; first car owned. Content of experience is measured by satisfaction with previous car and dealer. A direct measure of prior information is provided by responses to questions about whether respondents knew which manufacturer or dealer they wanted to buy from before search. These questions had the following response scale: definitely yes = 1, 0 otherwise. Time costs are measured by the respondent's hourly wage, a crude measure that may also be related to gains to search (high wage people have more alternatives).[4] Education is measured as years of school completed. Our age measure consisted of dummy variables for the following categories: age under 30 (AGE 25), age 31-40 (AGE 35), the omitted category is age 41-50, age 51-60 (AGE 55), age 61 and above (AGE 65). In addition to these measures we included other demographics, such as gender, marital status (married or not), employment status.

Our measure of Internet use was a self-report of time spent searching for information about cars on the Internet. We also had an alternative measure of whether respondents consulted the Internet for information about cars; the two measures provided consistent results in all but a few cases. We also asked respondents to check a list of types of information that they obtained on the Internet, and to provide a list of sources that they consulted on the Internet.

Descriptive Results

Since little is known about the characteristics of those who use the Internet in search for automobiles, a description of our sample responses is of interest in itself. Table 1 presents the

[4] The wage for those who do not work had to be imputed. We ran a first-stage probit analysis of those who work and those who do not work, and computed a selectivity bias correction factor from this analysis. However, this correction factor was statistically insignificant in a second stage regression of log wage on its determinants for the working sample. Accordingly we used a standard regression of log wage on male, married, education, number of

basic characteristics of our samples. To make monetary measures for the two samples comparable, monetary measures for 1989 were inflated to 1999 values using the Consumer Price Index. Possibly reflecting improvements in the industry over the past 10 years, the 1999 sample tended to more satisfied with their previous car and dealer. Possibly as a result of this, respondents tended to consider fewer models and dealers in 1999 and to search less. More of the 1999 respondents were female, and, because of differences in classifications of vehicles by our mailing list supplier, the 1999 sample includes more minivans. Compared to others in the 1999 sample, Internet users tended to search more in general, and to consider more models and dealers. They tended to be less satisfied with the dealer and less likely to know which dealer to buy from. They were younger, more educated, and had a higher wage. As table 1 indicates, median amounts of search are considerably below the means for the various samples, which reflects the fact that data on search are skewed.

(Table 1 about here)

Table 2 presents a description of Internet use by our 1999 sample. Slightly less that 40 percent of the respondents reported using the Internet in their search. This is very close to the proportion of Internet users for the U.S. as a whole reported in J.D. Power (2000). Virtually all of our respondents who reported using the Internet reported using this medium for price search, about two-thirds of the Internet users reported using the Internet for performance and reliability information. Slightly less than a third of the Internet users reported obtaining a referral to a dealer.

Responses to the sources consulted question were unaided responses to an open-ended question. By far the most common mention of sources consulted was the manufacturer page; the

---

children, age dummies and household income for the working sample to develop an imputed wage for those who do not work.

proportion consulting this was roughly consistent with the proportion seeking performance and reliability information. In fact approximately 81 percent of those who contacted the manufacturer home page reported seeking either performance or reliability information on the Internet. Evidently the manufacturer home page has become a major information source for car buyers. Other sources mentioned frequently were *Consumer Reports*, various sources of price information, and dealerships.

(Table 2 about here)

Table 3 presents a descriptive logistic regression contrasting Internet users and non users. This gives very similar results to a Tobit relationship of share of use of the Internet as a function of predictors that is consistent with our theoretical model, and that will be presented later. The difference between the two analyses is that the one in Table 3 does not control for differences between respondents in overall incentives to search (differences in g, w, S, A), while the later one does control for these because share is used as the dependent measure. Table 3 indicates that use of the Internet has a strong inverse relation to age, a strong positive relation to education and sticker price. The latter effect may be partly due to the greater latitude to negotiate price on expensive cars; alternatively it may simply be due to collinearity between price and the set of car type dummy variables. Internet use is also inversely related to prior knowledge of what manufacturer or dealer to buy from, and to satisfaction with the dealer. The negative signs on various car types are relative to the null category of small car, and may, as noted above, to some extent be an artifact of collinearity with the sticker price variable.

(Table 3 about here)

Table 4 presents logistic regressions of type of information sought on the Internet on the same variables as included in the Internet choice regression. While the results are similar to those

for the choice regression, there are some notable exceptions. Seeking price information is more closely related to the wage, a measure of time costs, suggesting that the Internet may be used as a time saving way to obtain price information. Using the Internet for price information is less related to lack of prior knowledge of which manufacturer to buy from, suggesting that this information is used to gather information for the upcoming negotiation once a car has been decided upon. Use of the Internet for performance and reliability information, on the other hand, is strongly related to not knowing which manufacturer to buy from. It is also significantly related to dissatisfaction with the previous dealer, suggesting that users may turn to the Internet for performance and reliability information because they do not see the dealer as a satisfactory source of this information. Finally, use of the Internet for performance and reliability information is signifcantly related to buying an upscale sedan (such as a Lexus 300), a type of car for which this type of information might be particularly salient. On the other hand, using the Internet for reliability information is not related to sticker price, suggesting that those who are concerned with this information may buy less expensive cars. Finally, possibly because it is less common, Internet referral to a dealer is less well explained than other information types. One finding is a possible negative relationship to number of cars bought in the past 10 years, suggesting that less experienced buyers are more likely to use the Internet for a referral.

(Table 4 about here)

Results for Share of Use of Sources

Internet Share

Table 5 presents analyses the share of total hours of search of the Internet as a function of predictor variables. The first run uses all of the predictors employed in the logistic regression in Table 3, the second eliminates the dummies for car types, while the third contains only the

variables that have an a priori reason to affect the relative productivity of the Internet. The analyses document the strong effects of age and education, which are not surprising, but worth verifying. Other results are the negative effects of decision to buy from a given manufacturer prior to the search (only significant at .08), and satisfaction with the previous dealer, and a positive effect of wage.

(Table 5 about here)

A Likelihood ratio test of Model 2 vs. Model 1 yielded a Chi-Square of 9.02 with 8 degrees of freedom, not significant at .05. Thus the car type searched for has no significant effect on the attractiveness of the Internet as a source. Sticker price becomes less significant when these variables are removed, indicating effects of collinearity between type and price. A likelihood ratio test of Model 3 vs. Model 1 had a Chi-Square value of 21.96 with 18 degrees of freedom, and a similar test of Model 3 vs. Model 2 had a Chi-Square value of 12.94 with 10 degrees of freedom. Neither of these results is significant. Model 3, which is restricted to variables that had some a priori rationale for inclusion, can be preferred. Our evidence indicates that variables related to gains to search and prior information affect the productivity of the Internet in predictable ways.

The IIA Assumption

Our empirical model rests on the assumption that the presence of the Internet does not affect the relative share of the remaining sources: the Internet draws share equally from all of the other sources. The study by Klein and Ford (2001) suggests that this is a reasonable assumption. As noted above, the study by Klein and Ford (2001) indicates that there are three distinct dimensions of Internet use, with their manufacturer/dealer dimension of Internet use corresponding to other manufacturer/dealer-related sources, their buying services and other third

parties dimension corresponding to other non-advocate sources, and bulletin board/chat sources corresponding to friends/relatives. Their evidence indicates that the Internet provides information that is provided by all three categories of sources, and thus substitutes for all three categories. Consequently the proportional draw hypothesis is a reasonable expectation.

One test of our proportional draw assumption about substitution of the Internet for other sources is presented in Table 6. Table 6 shows that the Internet accounted for about 7.3 percent of the reported search time of an average respondent; since 60 percent of the sample did not use the Internet, the average among users is much higher. Average shares of the other 3 sources declined correspondingly between 1989-99. However, if one excludes the Internet in 1999 (estimates the average relative to *A*), average shares of the other 3 sources do not change significantly.

(Table 6 about here)

If the proportional draw assumption is violated, shares of the remaining 3 sources excluding the Internet (using *A* as the base) should change with Internet use. To test this we regressed the shares of the 3 sources excluding the Internet on the Internet share. The results in Table 7 show that Internet share was never significant in these regressions, indicating that our independence assumption is reasonable. This does not mean that the Internet has no effect on the use of other sources; it merely means that the Internet appears to draw proportionally from the original sources. For example, if a consumer originally spent all search time with the dealer, and now spends 20 percent with the Internet, the dealer's share will fall to 80 percent. In fact, because the dealer/manufacturer source is the most extensively used, our results imply that the Internet will have the largest *absolute* effect on this source.[5]

[5] Most of the reported time with this source is directly related to activities at the dealer, such as test drives, time with salesmen.

(Table 7 about here)

Determinants of Shares of Other Sources

Tables 8-9 present analysis of the determinants of use of the other three sources if the Internet share is removed. Table 8 presents a tobit analysis of the friend share on pooled 1989-99 data (results for the separate samples were very similar). As discussed above, we expect the share of friends/relatives in the search process to be inversely related to number of cars bought in the past 10 years, age, education and income, positively related to being female, single and buying one's first car. For the most part these results are borne out in Table 8. Notable exceptions are that first car has the wrong sign, and gender and wage are insignificant. While satisfaction with dealer and satisfaction with manufacturer were also included in Table 8, neither variable was significant. In general our results indicate that friend is a relatively productive source of information for younger, single, less experienced and educated automobile buyers.

(Table 8 about here)

Table 8 also presents a tobit analysis of the non-advocate share of search time excluding the Internet on pooled 1989-99 data (results for the separate samples were very similar). This share is inversely related to prior knowledge of the manufacturer to buy from, satisfaction with the dealer, and to experience as measured by number bought in the past 10 years. It is positively related to education and being male, but not to price paid, wage, or age. Evidently more educated consumers of either gender who are younger and need information are likely to consult the Internet, while more educated consumers who are older and male are likely to consult print sources for similar information. In contrast less educated consumers who are younger are likely to consult friends or relatives.

Since all consumers reported positive use of the dealer, we ran regressions of dealer share excluding the Internet on the pooled 1989-99 sample instead of tobit analysis. The results are presented in Table 9. Possibly because they do not utilize other sources, less educated, more experienced buyers tend to concentrate their search at the dealer. A result that we have difficulty explaining is that the dealer share is significantly related to the number of children at home, and possibly to being female (significant at .09). Possibly female buyers with children at home do not have time to search from other sources.

(Table 9 about here)

Determinants of Total Search

Let $A_{ck} = A_{0k}\left(1/\left(1 - s_{nk}\right)\right)$, where s_{nk} is Internet share in search. Then Equation 12 can be written as:

$$(14.)\quad \ln T_k = \frac{\left(\ln\left(A_{ck}\right) + \ln g_k - S_k - \ln w_k - \left(A_{ck}\right)\left(\sum_j s_{1j} \ln s_{1j} + s_{nk} \ln s_{nk}\right)\right)}{A_{ck} + 1},$$

By substituting shares and appropriate predictors of *A*, *g*, *S*, *w*, we can estimate Equation 14 by nonlinear regression. We hypothesize that A_{ok}, which is related to efficiency, is a function of education and marital status (the spouse can lower the need for the respondent to obtain information):

(15.) $A_{ok} = a_0 + a_1$*Education + a_2*Marital Status.

We hypothesize that *g* is related to price, since more expensive cars provide more scope for negotiation and involve more complex tradeoffs. Since the respondent's wage may be partly related to gains to search we relax the constraint that the coefficient of ln *w* should be one. Other terms in Equation 16 below are measures of prior information.

(16.) $\ln g_k - S_k - \ln w_k = b_0 + b_1$*Know Manufacturer+$b_2$*Know Dealer+$b_3$*First Car +$b_4$*Satisfied Car+$b_5$*Satisfied Dealer+$b_6$*No. New 10 Yr.+$b_7$*Sticker Price+ b_8*Log of wage+b_9*Age25+b_{10}*Age35+b_{11}*Age55+b_{12}*Age65

Substituting Equations 15 and 16 into Equation 14, and also incorporating the various shares (which are exogenously determined), Equation 14 was estimated on pooled data for 1989 and 1999.[6] Because s_{nk} is incorporated into Equation 14 this estimate incorporates the effect of the Internet on total search.

The results of the estimation are presented in Table 10. Table 10 shows that education and marital status are insignificant, and the hypothesis that A_{ok} is a constant cannot be rejected. If correct, the implication is that, controlling for prior information and differences in the relative efficiency of sources, the ability to search for automobiles is the same across consumers.

(Table 10 about here)

What appears to determine differences in search are gains to search and prior information. The coefficient of price is significantly positive, and the coefficient of ln *w* is greater than –1, indicating that this variable is related to gains to search as well as time costs. For the most part, factors related to prior information have the expected negative impact on search. A prior decision to buy from a particular manufacturer or dealer, satisfaction with the previous dealer, and age are all inversely related to time spent searching.

Gains from the Internet

The estimates of the share equations and the expression for total search in Table 10 provide all of the parameter estimates needed to determine gains for the Internet. Using these parameter estimates Equation 1 was computed for selected groups on the 1999 sample under two

[6] Estimating equation 14 on 1989 data alone creates a problem of collinearity between a_o in Equation 15 and b_o in Equation 16. Pooling the data (or estimating on the 1999 sample) mitigates this problem because Internet shares are introduced into the expression for *A*.

alternatives assumptions: Internet was present as it was in 1999; Internet was eliminated as a source.

Table 11 summarizes the results of this exercise for the sample as a whole and for 5 subgroups. The various shares with the Internet present are actual shares in the data. The table shows that the Internet share of time spent searching is the highest for the Internet users, and that this source is second in popularity for this group. Most of this added share appears to be drawn from the dealer. In contrast the Internet share is much lower for less educated and older groups, who tend to devote relatively more of their time to the dealer/manufacturer sources. The younger, less educated, consumer makes the most use of the interpersonal sources (friend), least use of the non-advocate sources.

(Table 11 about here)

The remaining entries in Table 11 are estimates drawn from our analysis of time spent, and the proportional draw assumption. The estimates of overall time spent come from Model 1 in Table 10, and assume a lognormal error. Because they are based on predicted values that remove the influence of outliers, these estimates are more consistent with the medians reported in Table 1 than the means.

The estimates of total time spent indicate an increase in time spent if the Internet were removed as a source; one efficiency provided by the Internet is less time required to gather information. While Internet users and groups that are prone to use the Internet tend to use more time than others, our estimates indicate that they would use even more if the Internet were absent. This is because these younger groups have the least prior information, and therefore more to gain from search. All groups are seen to gather more information, as revealed by F, than they would if the Internet were absent. As must be the case, the Internet users would sacrifice the

most information if the Internet were removed. If the Internet were removed, our estimates indicate that Internet users, and other groups that are likely to use the Internet, would substantially increase their time with dealer/manufacturer sources.

In terms of Equation 1, unrealized gains are computed as the difference between *g* under perfect information and *g* given the information obtained from the search. This is equal to $ge^{-(S+F)}$. As they must be, the unrealized gains are smaller when the Internet is present. Internet users would leave \$97.39 – \$76.65 = \$17.74 more on the table if the Internet were not there, if our estimates are correct. The total gain from having the Internet is computed as the difference B1 – B0, where B1 is net gain from the search computed from Equation 1 with Internet present, B0 is net gain from the search computed from Equation 1 with the Internet absent. This gain, which includes both the value of better buys and savings in search costs, is actually highest for the young, educated group, which has the most to gain from search (especially those who are relatively dissatisfied with their dealer). For the latter group the gain is estimated to be \$38.50. In contrast the estimated average gain from having the Internet is only \$2.71 for the over 50 group, and \$7.75 for the less educated younger group. The welfare gains from the Internet are obviously skewed toward the younger, more educated, consumers who have more to gain from search.

Conclusions

We have developed and estimated a model in which total search depends on potential gains, prior information, productivity at obtaining information and time costs. In this framework, the addition of the Internet must increase productivity (or at worst leave it unchanged). While this increase in productivity has ambiguous effect on total search effort - it lowers cost of information, but decreases the time needed to obtain a given amount of information – our empirical evidence in the preceding section indicates that the Internet leads to reduced search.

This is even though Internet users tend to search the most. Our evidence indicates that they do so because they have the most to gain from search and therefore benefit the most from the presence of this source.

Internet use in search for cars depends heavily on young age, high education, high income. About 40% of our 1999 sample used Internet, and almost all of these reported using it for price information. Referrals to dealer, which are reported by about 30 percent of Internet users, tend to be used by less experienced buyers. Internet users are also the ones likely to have the most to gain from search, and they use other sources as well. In particular, they:

- Buy more expensive cars – high g.
- Are young, less likely to make a choice prior to search, likely to be less satisfied with previous dealer, so have less prior information, or lower S.
- Are more educated – efficient at processing information (and using Internet).

According to our estimates, the Internet affects the use of other sources in approximately the same proportion. However, it has the biggest absolute effect on use of dealer/manufacturer sources (dealer accounts for the biggest proportion of the time in this category) since this is the highest share source. Our analysis further indicates that total search across all sources increases with gains to search, as proxied by sticker price, but decreases with costs of search, as measured by log of wage, and decreases with variables related to prior information, including prior knowledge of what manufacturer or dealer to purchase from, satisfaction with the previous dealer, and age.

We used our estimates of shares and total search to develop simulations of behavior in the presence or absence of the Internet. Our analysis indicates that Internet users spent more time overall in search because they had the greatest incentives to search, and relatively low amounts

of prior information. According to our estimates, the availability of the Internet allowed the users to spend less time searching than otherwise, and led to both savings in search costs and better buys. Corroborating evidence that Internet users – at least those who get price referrals – do get better buys is presented by Morton, Zettelmeyer and Risso (2000). Those who gained the most from the Internet were younger, well educated, consumers, while older consumers stood to gain very little. Aside from their being less likely to use the Internet, this group had the most prior information, the least to gain from search, and therefore searched the least.

The simulation results presented in Table 11 indicate that the presence of the Internet leads to substantial reductions in time with dealer/manufacturer sources. Since our data indicate that most of this is time spent at dealerships that is likely to involve a sales person, the Internet appears to provide substantial efficiency gains to dealers as well as consumers. Dealers save in selling time if consumers come in armed with information that they would not have obtained if the Internet were not present.[7] Thus our evidence suggests that the Internet provides efficiency gains not only to consumers but also to dealers. This, in turn, is likely to have second-order effects on prices (one might expect buyers and dealers to share the gain), and on the structure of automobile retailing.

Our estimates of gains to search are those implied by our model, which are predicated on consumers attempting to maximize the net benefit of search. This is analogous to the profit maximization or cost minimization assumptions commonly made in productivity calculations, and evidence presented in Ratchford and Srinivasan (1993) indicates that this assumption is a reasonable approximation for automobile buyers. However, it would be useful to have more direct evidence about the impact of the Internet on outcomes obtained by consumers. Morton, Zettelmeyer and Risso (2000) do present evidence about prices obtained by those using a referral

service. However, it is difficult to determine how the users of the service might have fared in the absence of the Internet – they may have been able to get low prices in any case. Our data may allow us to determine the impact of using the Internet on actual outcomes, which is a project that we hope to undertake in the near future.[8]

[7] Morton, Zettelmeyer and Risso (2000) make a similar point.

[8] The helpful comments of participants in seminars at the University of Houston, University of Connecticut, University of Maryland, and University of Texas at Dallas are gratefully acknowledged.

Table 1
Description of Data

	1989	1999	Diff. 99-89	Non-internet	Internet	Difference
Label	Mean	Mean	Sig.at .05	Mean	Mean	Sig.at .05
Models Seriously Considered	2.91	2.72	S	2.52	3.04	s
Dealers Seriously Considered	2.50	2.27	S	2.14	2.49	s
Know Manufacturer To Buy From	0.47	0.48	N	0.55	0.37	ns
Know Specific Dealer To Buy From	0.27	0.27	N	0.34	0.16	0.01
First Car Ever Owned	0.02	0.04	S	0.03	0.05	s
Satisfied With The Dealer	5.17	5.50	s	5.64	5.27	s
Satisfied With Previous Car	5.40	5.84	s	5.89	5.74	ns
Number New Cars Purchased 10 Yr.	1.91	1.54	s	1.58	1.48	ns
Sticker Price Of The New Car (1999$)	19675	21319	s	20936	21926	s
Male	0.67	0.52	s	0.50	0.54	ns
Married	0.69	0.66	ns	0.64	0.70	s
Years Of Education-Respondent	14.73	15.04	s	14.47	15.94	s
Number Of Children	0.83	0.84	ns	0.67	1.11	s
Number Of Cars In The Household	2.09	1.97	s	1.87	2.12	s
Years Lived In The Area	23.27	25.70	s	29.41	19.98	s
Employment Status-Respondent	0.75	0.70	s	0.60	0.85	s
Age In Years	46.34	48.28	s	52.62	41.41	s
Dollar Household Income (1999$)	59846	59583	ns	52205	71319	s
Dollar Wage – Respondent (1999$)[a]	23.72	21.98	s	19.65	25.72	s
Search Time – Internet		1.22	na	0.00	3.10	na
Total Search Time	18.61	15.58	s	13.47	18.81	s
Median of total search time[b]	13.92	11.50	s	9.58	15.00	s
Bought Small Car	0.21	0.27	s	0.26	0.29	ns
Bought Family Sedan	0.51	0.36	s	0.38	0.32	ns
Bought Sporty Car	0.07	0.06	ns	0.06	0.06	ns
Bought Large Sedan	0.19	0.07	s	0.08	0.05	s
Bought Upscale Sedan		0.01	na	0.01	0.03	s
Bought Luxury Sedan		0.02	na	0.03	0.01	s
Bought Station Wagon		0.04	na	0.04	0.04	ns
Bought Minivans	0.01	0.17	s	0.14	0.20	S
Bought Other	0.00	0.01	ns	0.01	0.01	ns

a Includes imputed wage for those who do not work. Imputed wage is value predicted from a regression relation between wages and demographics for those who work (see footnote 3).
b Cell entries for this row are medians and not means.

Table 2
Description of Internet Use

Variable		Pct.
Used Internet		38.95%
Internet price Information		37.58%
Internet performance information		28.56%
Internet reliability Information		25.40%
Internet referral to dealer		11.85%
N		886
Sources Consulted on Internet		
	No. Mentions	Pct.
Manufacturer page	108	29.75%
Edmonds	47	12.95%
Auto by tel	43	11.85%
Kelly Blue Book	31	8.54%
Consumer Report	21	5.79%
Yahoo car guides	16	4.41%
Dealership page	14	3.86%
Carpoint	12	3.31%
AOL Auto Place	11	3.03%
Priceline	3	0.83%
Other	57	15.70%
Total Mentions	363	100.00%
Total Users	345	
Mentions/User	1.05	

Table 3
Logistic Regression – Internet Users vs. Non-Users

Variable	Parameter Estimate	Chi-Square	Pr > ChiSq
Intercept	-3.1116	16.11	0.00
Know Manufacturer To Buy From	-0.3686	4.04	0.04
Know Dealer To Buy From	-0.3694	2.76	0.10
First Car Ever Owned	-0.2894	0.50	0.48
Satisfied With Dealer	-0.1055	3.16	0.08
Satisfied With Previous Car	0.0457	0.46	0.50
No. New Cars Last 10 Yr.	-0.0309	0.19	0.67
Sticker Price Of The New Carx10^{-4}	0.6000	10.07	0.00
Male	0.0911	0.26	0.61
Married	0.2505	1.49	0.22
Years Of Education-Respondent	0.0994	7.56	0.01
No. Children In Household	0.0361	0.18	0.67
Years Lived In The Area	-0.0039	0.51	0.48
Employment Status-Respondent	0.1637	0.36	0.55
Ln(Wage)	0.2889	2.31	0.13
Age25	0.9164	9.18	0.00
Age35	0.3451	2.10	0.15
Age55	-0.6905	6.83	0.01
Age65	-1.4976	19.64	0.00
Bought Family Sedan	-0.3995	3.15	0.08
Bought Sporty Car	-0.5353	1.81	0.18
Bought Large Sedan	-0.4311	1.10	0.29
Bought Upscale Sedan	0.7162	0.73	0.39
Bought Luxury Sedan	-2.6357	7.16	0.01
Bought Station Wagon	-0.3491	0.63	0.43
Bought Minivans	-0.4402	2.13	0.14
Bought Other	-1.5593	3.84	0.05
Rho-squared	.32		

Table 4
Logistic Regression: Choice of Information Type from Internet

	Price Information		Performance Information		Reliability Information		Referral to Dealer	
Variable	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq
Intercept	-3.054	0.00	-3.934	0.00	-2.828	0.00	-1.829	0.07
Know Manuf.	-0.249	0.17	-0.544	0.00	-0.378	0.05	-0.357	0.16
Know Dealer	-0.326	0.14	-0.318	0.19	-0.341	0.16	-0.474	0.17
First Car	-0.364	0.37	-0.377	0.38	-0.375	0.38	-0.573	0.32
Satisfied Dealer	-0.075	0.20	-0.142	0.02	-0.147	0.01	-0.005	0.95
Satisfied Car	-0.030	0.65	0.103	0.14	0.089	0.21	-0.134	0.12
No. New 10 Yr.	-0.075	0.29	-0.037	0.62	-0.022	0.77	-0.185	0.07
Sticker Price $x10^{-4}$	0.510	0.00	0.370	0.03	0.120	0.44	0.044	0.84
Male	0.115	0.51	0.085	0.65	-0.099	0.59	0.256	0.29
Married	0.125	0.54	-0.094	0.66	-0.049	0.82	0.043	0.87
Education	0.096	0.01	0.143	0.00	0.144	0.00	0.120	0.01
Children	0.022	0.80	0.069	0.44	-0.010	0.91	-0.083	0.49
Years In Area	-0.004	0.41	-0.001	0.91	-0.002	0.76	-0.002	0.79
Employment	0.176	0.51	0.084	0.77	-0.061	0.83	-0.468	0.20
Ln(Wage)	0.412	0.03	0.244	0.20	0.091	0.64	-0.005	0.99
Age25	0.866	0.00	0.806	0.01	0.611	0.04	-0.010	0.98
Age35	0.403	0.09	0.257	0.29	0.160	0.51	0.187	0.55
Age55	-0.656	0.01	-0.682	0.02	-0.726	0.01	-0.634	0.12
Age65	-1.124	0.00	-1.055	0.00	-1.107	0.00	-0.863	0.07
Family Sedan	-0.325	0.14	-0.006	0.98	0.078	0.73	-0.343	0.24
Sporty Car	-0.302	0.44	0.220	0.58	0.067	0.87	0.058	0.91
Large Sedan	-0.363	0.36	-0.349	0.43	-0.598	0.22	-0.966	0.15
Upscale Sedan	0.809	0.32	1.684	0.04	1.618	0.03	-0.207	0.81
Luxury Sedan	-3.142	0.01	-1.343	0.15	-0.540	0.56	-13.985	0.99
Station Wagon	0.170	0.69	0.205	0.64	0.361	0.40	-0.345	0.57
Minivans	-0.335	0.26	-0.206	0.49	-0.046	0.88	-0.147	0.70
Other	-1.357	0.09	-0.637	0.41	-0.244	0.74	0.778	0.31
Rho-squared	0.29		0.25		0.19		0.09	

Table 5
Tobit Analysis of Share of Internet in Total Search

	Model 1		Model 2		Model 3	
Variable	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq
Intercept	-0.4512	0.000	-0.4287	0.000	-0.3780	0.000
Know Manufacturer	-0.0296	0.186	-0.0312	0.164	-0.0357	0.077
Know Dealer	-0.0155	0.570	-0.0160	0.556		
First Car Owned	-0.0346	0.498	-0.0300	0.557		
Satisfied With Dealer	-0.0150	0.035	-0.0169	0.017	-0.0154	0.0229
Satisfied With Previous Car	0.0083	0.312	0.0091	0.271		
No. New Last 10 Yr.	-0.0032	0.717	-0.0025	0.778		
Sticker Price x 10^{-4}	0.0427	0.020	0.0176	0.248		
Male	0.0150	0.491	0.0125	0.566		
Married	0.0291	0.245	0.0229	0.352		
Education	0.0176	0.000	0.0178	0.000	0.0181	0.000
No. Children	-0.0027	0.802	-0.0037	0.719		
Years In Area	-0.0007	0.318	-0.0008	0.257		
Employment Status	0.0306	0.349	0.0354	0.277		
Ln(Wage)	0.0367	0.112	0.0381	0.098	0.0557	0.005
Age25	0.0969	0.007	0.0946	0.009	0.0944	0.005
Age35	0.0596	0.038	0.0568	0.046	0.0625	0.024
Age55	-0.0702	0.037	-0.0753	0.025	-0.0776	0.014
Age65	-0.1332	0.001	-0.1415	0.001	-0.1682	0.000
Bought Family Sedan	-0.0448	0.090				
Bought Sporty Car	-0.0492	0.323				
Bought Large Sedan	-0.0631	0.197				
Bought Upscale Sedan	-0.0055	0.946				
Bought Luxury Sedan	-0.2846	0.011				
Bought Station Wagon	-0.0540	0.312				
Bought Minivans	-0.0479	0.168				
Bought Other	-0.0430	0.627				
Scale	0.2309		0.2327		0.2344	
Log Likelihood	-223.93		-228.44		-234.91	
pseudo-R sq	0.23		0.21		0.20	
N	813		813		819	

Table 6
Share of Internet in Total Search Time

Variable	1989 Mean	1999 Mean	1999 Excl. Internet	Excl. Internet Difference 99-89
Share Friend	0.099	0.090	0.097	ns
Share Nonadv	0.118	0.096	0.104	ns
Share Dealer	0.783	0.741	0.799	ns
Share Internet	0.000	0.073		

Table 7
Regressions of Share of Source Excluding Internet – Share Predicted from 1989 Data

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|
Friend share – predicted share				
Intercept	-0.0022	0.0049	-0.44	0.66
Internet share	-0.0295	0.0330	-0.89	0.37
R-Squared	0.0009			
Nonadvocate share – predicted share				
Intercept	-0.0177	0.0052	-3.43	0.00
Internet share	0.0500	0.0349	1.43	0.15
R-Squared	0.0025			
Dealer share – predicted share				
Intercept	0.0206	0.0069	2.98	0.00
Internet share	-0.0294	0.0466	-0.63	0.53
R-Squared	0.0005			
Sample Size	807			

Table 8
Tobit Analysis of Share of Friend and Non-Advocate (Excluding Internet)
Pooled 1989-99 Data

	Share of Friend			Share of Non-advocate		
Variable	Estimate	Chi-Square	Pr > ChiSq	Estimate	Chi-Square	Pr > ChiSq
Intercept	0.2371	47.08	0.00	0.0064	0.03	0.87
Know Manufacturer	0.0050	0.31	0.58	-0.0404	15.80	0.00
Know Dealer	-0.0135	1.73	0.19	-0.0039	0.11	0.74
First Car	-0.0449	3.42	0.06	-0.0172	0.38	0.54
Satisfied Dealer	0.0030	1.22	0.27	-0.0074	5.97	0.01
Satisfied Manufacturer	-0.0017	0.37	0.54	-0.0047	2.17	0.14
No. New 10 Yr.	-0.0168	29.43	0.00	-0.0063	3.38	0.07
Sticker Price x 10^{-4}	-0.0092	2.05	0.15	0.0039	0.30	0.59
Male	-0.0055	0.37	0.54	0.0356	12.08	0.00
Married	-0.0293	9.15	0.00	0.0093	0.71	0.40
Education	-0.0022	1.84	0.17	0.0071	14.82	0.00
No. Children	-0.0025	0.33	0.57	-0.0124	6.29	0.01
Years In Area	-0.0002	0.56	0.46	-0.0005	3.21	0.07
Employment Status	0.0110	0.71	0.40	0.0097	0.44	0.51
Log of wage	-0.0225	5.76	0.02	0.0128	1.50	0.22
Age25	0.0167	1.34	0.25	-0.0094	0.32	0.57
Age35	0.0189	2.53	0.11	0.0065	0.23	0.63
Age55	-0.0299	5.05	0.02	0.0099	0.43	0.51
Age65	-0.0309	3.78	0.05	-0.0097	0.30	0.59
Scale	0.1501			0.1693		
Pseudo R-sq.	0.11			0.07		
Pooling Test		35.72	> .01		30.18	> .05

Table 9
Regressions of Dealer Share (excluding Internet) Pooled 1989-99 Data[a]

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.7323	0.0358	20.44	0.00
Know Manufacturer	0.0142	0.0094	1.51	0.13
Know Dealer	-0.0050	0.0106	-0.47	0.64
First Car	0.0381	0.0258	1.48	0.14
Satisfied Dealer	0.0029	0.0029	1.02	0.31
Satisfied Manufacturer	0.0043	0.0030	1.46	0.15
No. New 10 Yr.	0.0156	0.0032	4.87	0.00
Sticker Price x 10^{-4}	0.0055	0.0066	0.84	0.40
Male	-0.0162	0.0095	-1.71	0.09
Married	0.0150	0.0102	1.47	0.14
Education	-0.0032	0.0017	-1.83	0.07
No. Children	0.0120	0.0046	2.61	0.01
Years In Area	0.0004	0.0003	1.42	0.16
Employment Status	-0.0133	0.0135	-0.99	0.32
Log of wage	0.0024	0.0097	0.25	0.80
Age25	0.0018	0.0155	0.12	0.91
Age35	-0.0141	0.0127	-1.12	0.26
Age55	0.0101	0.0139	0.73	0.47
Age65	0.0179	0.0165	1.09	0.28
R-Sq.	0.05			

a. Pooling test for equality of coefficients across samples has F = 0.58 ns at .10

Table 10
Regressions of Log of Total Search on Determinants

	Model 1			Model 2		
Parameter	Estimate	Std Error	T	Estimate	Std Error	T
Determinants of A_{ok}						
Intercept	3.0200	0.9617	3.14	3.1367	0.9873	3.18
Education				-0.0196	0.0196	-1.00
Marital Status				0.0856	0.1068	0.80
Determinats of *g,S*						
Intercept	9.7651	2.2059	4.43	9.6956	2.1184	4.58
Know Manufacturer	-1.2793	0.3484	-3.67	-1.2422	0.3298	-3.77
Know Dealer	-1.4775	0.4054	-3.64	-1.4270	0.3823	-3.73
First Car Ever Owned	0.0959	0.5108	0.19	0.0623	0.4933	0.13
Satisfied With Car	-0.0106	0.0579	-0.18	-0.0080	0.0563	-0.14
Satisfied With Dealer	-0.1308	0.0640	-2.04	-0.1322	0.0623	-2.12
No. New 10 Yr.	-0.0550	0.0618	-0.89	-0.0440	0.0604	-0.73
Sticker Price x 10^{-4}	0.4552	0.1698	2.68	0.4585	0.1654	2.77
Log of wage	-0.4332	0.1877	-2.31	-0.4959	0.2091	-2.37
Age25	0.3885	0.2999	1.30	0.3059	0.2905	1.05
Age35	-0.0287	0.2454	-0.12	-0.0399	0.2397	-0.17
Age55	-0.5189	0.2866	-1.81	-0.5029	0.2765	-1.82
Age65	-0.8385	0.3052	-2.75	-0.8077	0.2915	-2.77
R-squared	.24			.24		

Table 11
Comparison between Internet Present and Absent – Selected Groups 1999 Sample

Variable	Overall Mean	Internet Users Mean	Age < 40 College Grad Mean	Age < 40 HS or Less Mean	Age 50+ Mean	Age < 40 College Dealer Sat < 6 Mean
Internet Present						
Share Friend	0.09	0.07	0.09	0.14	0.08	0.07
Share Non-advocate	0.10	0.11	0.11	0.08	0.10	0.10
Share Dealer	0.74	0.63	0.65	0.71	0.78	0.65
Share Internet	0.07	0.19	0.15	0.06	0.04	0.18
Overall Time Spent (hr.)	12.33	14.09	14.53	14.62	10.50	14.18
Time Friend (hr.)	1.11	0.99	1.31	2.05	0.84	0.99
Time Non-advocate (hr.)	1.23	1.55	1.60	1.20	1.05	1.42
Time Dealer (hr.)	9.12	8.88	9.44	10.45	8.19	9.22
Time Internet (hr.)	0.86	2.68	2.18	0.92	0.42	2.55
F - $\sum a_i \ln t_i$	5.34	5.72	5.97	5.93	4.84	6.07
S – Prior Information	2.17	1.66	1.36	1.42	2.86	1.33
Unrealized Gains ($)	63.89	79.65	78.91	54.92	53.00	77.52
Internet Absent						
Share Friend	0.10	0.09	0.11	0.15	0.08	0.10
Share Non-advocate	0.11	0.13	0.13	0.09	0.10	0.11
Share Dealer	0.80	0.78	0.76	0.77	0.81	0.79
Overall Time Spent (hr.)	12.40	14.29	14.97	14.60	10.46	14.81
Time Friend (hr.)	1.20	1.29	1.65	2.17	0.88	1.48
Time Non-advocate (hr.)	1.31	1.86	1.95	1.25	1.08	1.63
Time Dealer (hr.)	9.89	11.15	11.38	11.17	8.50	11.70
F - $\sum a_i \ln t_i$	5.25	5.47	5.93	5.86	4.80	5.76
Unrealized Gains ($)	70.32	97.39	96.40	59.89	55.45	96.94
Comparison						
Gain from Having Internet ($)	10.20	28.17	33.33	7.75	2.71	38.50
N	869	319	161	31	401	64

References

Alba, Joseph, John Lynch, Barton Weitz, Chris Janiszewski, Richard Lutz, Alan Sawyer, and Stacy Wood (1997), "Interactive Home Shopping: Consumer, Retailer and Manufacturer Incentives to Participate in Electronic Marketplaces," *Journal of Marketing*, 61, 3 (July), pp. 38-53.

Anonymous (2000), "Study Shows 300 Mil Worldwide Web Users," *Newsbytes News Network*, March 22.

Beatty, Sharon E. and Scott M. Smith (1987), "External Search Effort: An Investigation Across Several Product Categories," *Journal of Consumer Research*, 14 (June), 83-93.

Bakos, J. Yannis (1997), "Reducing Buyer Search Costs: Implications for Electronic Marketplaces," *Management Science*, 43, 12 (December), pp. 1613-1630.

Consoli, J. (2000), "Counterpoint: The Media Respond," *Mediaweek*, July 17, p. 7.

Dreazen, Y. (1999), "Online Ad Spending Is Expected to More Than Triple by Year 2004," *The Wall Street Journal*, p. B12

Furse, David H., Girish N. Punj and David W. Stewart (1984), "Typologies of Individual Search Strategies Among Purchasers of New Automobiles," *Journal of Consumer Research*, 10 (March), pp. 417-431.

Gablehouse, G. (2000), "Print Is Alive And Well: Study," *Mediaweek*, July 17, pp.6-7.

Hauser, J., G. Urban and B. Weinberg (1993), "How Consumers Allocate Their Time When Searching for Information," *Journal of Marketing Research*, 30 (November), pp. 452-466.

Hoffman D. L. (2000), "The Revolution Will Not Be Televised: Introduction to the Special Issue on Marketing Science and the Internet," *Marketing Science*, Vol. 19, No. 1(Winter), pp. 1-3.

Klein, Lisa R. and Gary T. Ford (2001), "Consumer Search for Information in the Digital Age: an Empirical Study of Pre-Purchase Search for Automobiles," working paper, Rice University.

J.D. Power and Associates (2000), *New Autoshopper.com Study*, Agoura Hills, CA.

Lal, Rajiv and Miklos Sarvary (1999), "When and How is the Internet Likely to Decrease Price Competition?" *Marketing Science*, 18 (4), pp. 485-503.

Lynch, John and Dan Ariely(2000), "Wine Online: Search Costs Affect Competition on Price, Quality and Distribution," *Marketing Science*, 19 (Winter), pp. 83-103.

Moorthy, K. Sridhar, Brian T. Ratchford and Debabrata Talukdar (1997), "Consumer Information Search Revisited: Theory and Empirical Analysis," *Journal of Consumer Research*, 23 (March), pp. 263-277.

Morton, Fiona Scott, Florian Zettelmeyer and Jorge Silva Risso (2000), "Internet Car Retailing," working paper, Yale University

Parker, B. J. and R. E. Plank (2000), "A Uses and Gratifications Perspective on the Internet as a New Information Source," *American Business Review*, June, pp. 43-49.

Ratchford, B. T., D. Talukdar and M.S. Lee (2001), "A Model of Consumer Choice of the Internet as an Information Source," *International Journal of Electronic Commerce*, 5 (Spring), pp. 7-21.

Ratchford, B.T. and N. Srinivasan (1993), "An Empirical Investigation of Returns to Search," *Marketing Science*, 12 (1).

Rutkowski, T. (1999), "Biannual Strategic Note: Internet Survey Reaches 56.2 Million Internet Host Level," *Center for Next Generation Internet NGI.ORG*. (August 12) http://www.ngi.org/pub/TrendsPR9908.txt